



PCT09

RAW SEQUENCE LISTING

DATE: 03/15/2002

PATENT APPLICATION: US/09/831,142A

TIME: 11:08:09

Input Set : A:\es.txt

Output Set: N:\CRF3\03152002\I831142A.raw

4 <110> APPLICANT: University of Wales College of Medicine
 6 <120> TITLE OF INVENTION: Protein and DNA coding therefor
 8 <130> FILE REFERENCE: PCT/GB99/03654
 C--> 10 <140> CURRENT APPLICATION NUMBER: US/09/831,142A
 C--> 11 <141> CURRENT FILING DATE: 2001-05-07
 13 <160> NUMBER OF SEQ ID NOS: 22
 15 <170> SOFTWARE: PatentIn Ver. 2.1
 17 <210> SEQ ID NO: 1
 18 <211> LENGTH: 870
 19 <212> TYPE: DNA
 20 <213> ORGANISM: Pholas dactylus
 22 <400> SEQUENCE: 1
 23 gaattcggca cgagtcggaa aagaacaaaa tggcttgtat cgttttcggt gctcttgtog 60
 25 ctctatgctt aatgcaaccg ggttcgggtg aggaagtaca atgcgcgatg aattggacac 120
 27 aagctaata ga atagtgttgc aacgtggact ggatgaccat ttcatctac gactatggcg 180
 29 ctcaagagca actgtacgaa gatcgggctt tggggctgtg tcggattgaa cgggccggcc 240
 31 caggtaccac aaaagccgtc tggattaact ggagtaacga cagcagtcac tgtgtaacaa 300
 33 gaaaaacaat cttcttcgag gttggtggag aaattgcccg gctagttagac tacagaccac 360
 35 aggaagacgg aactgagaaa actttttacaa gaaaattctc tagcaaaatg ccaggcactt 420
 37 acatgcttat ggacgtgtgc gctacaaggg acgctgatga taaatgcatc gaaggcaca 480
 39 ttgtggtgac agtcagggtg tccctatatg acgaagataa caatggtgta atggatgaag 540
 41 gtaagggtgat tccatctgag acaatcgagg atgatataca ggactgtggg ctcttagacc 600
 43 aagatgttga actcgtattat acgtggactc aaaaacgagt tgatctacca gacacagtag 660
 45 acgaggctga agacacaccg tcagaaactg gagaattctt ctggtagatc tatcagacta 720
 47 cttttatcag caggacaact ggtcgttacc agacacctat aacgtgtcct catcaataat 780
 50 gtgtaaaaca gaaataatcg atagaatatt gaaaataaaa tgttaataaa cactggttga 840
 52 aatatgaaaa aaaaaaaaaa aaactcagag 870
 56 <210> SEQ ID NO: 2
 57 <211> LENGTH: 816
 58 <212> TYPE: DNA
 59 <213> ORGANISM: Pholas dactylus
 61 <400> SEQUENCE: 2
 62 gaattcggca cgagggaata gaacaaaatg gcttgtatcg ttttcgttgc ttttgcgct 60
 64 ctatgcttaa tgcaaccggg ttccgggtgag gaagtacaat gcgcgatgaa ttggacacaa 120
 66 gctaataaat atgtgttcaa cgtggactgg atgaccattt tcatctacga ctatggcgct 180
 68 caagagcaac tgtacgagga tcgggctttg gggtgtgtc ggattgaacg ggccggccca 240
 70 ggtaccacaa aagccgtctg gattaactgg agtaacgaca cgcagtcacg tgtaacaaga 300
 72 aaaacaatct tcttcgaggt tgggtgagaa attgcccggc tagttgacta cagaccacag 360
 74 gaagacggaa ctgagaaaaa ttttacaaga aaattctcta gcaaaatgcc aggcacttac 420
 76 atgcttatgg acgtgtgcgc tacaagggac gctgatgata aatgcacatg aggcacaatt 480
 78 gtggtgacag tcagggtgtc cctatatgac gaagataaca atggtgtaat ggatgaagg 540
 80 aaggttattc catctgagac aatcgaggat gatatacagg actgtgggct cttagacca 600
 82 gatgttgaac tcgattatac gtggactcaa aacgagtgtg atctaccaga cacagtagac 660

ENTERED

RAW SEQUENCE LISTING

DATE: 03/15/2002

PATENT APPLICATION: US/09/831,142A

TIME: 11:08:09

Input Set : A:\es.txt

Output Set: N:\CRF3\03152002\I831142A.raw

```

84 gaggtgaag acacaccgtc agaaactgga gaattcttct ggtagatcta tcagaccact 720
86 tttatcagca ggacaactgg tcgttaccag acacctataa cgtgtcctca tcaataatgt 780
88 gtaaaacaga aataatcgat agaattattga aaataa 816
92 <210> SEQ ID NO: 3
93 <211> LENGTH: 852
94 <212> TYPE: DNA
95 <213> ORGANISM: Pholas dactylus
97 <400> SEQUENCE: 3
98 gtcggaag aacaaaatgg cttgtatcgt tttcgttgc cttgtcgtc tatgcttaat 60
101 gcaaccgggt tccggtgagg aagtacaatg cgcgatgaat tggacacaag ctaatgaata 120
103 tgtgttcaac gtggactgga tgaccatttt catctacgac tatggcgtc aagagcaact 180
105 gtacgaggat cgggctttgg ggctgtgtcg gattgaacgg gccggcccag gtaccacaaa 240
107 agccgtctgg attaactgga gtaacgacac gcagtcatgt gtaacaagaa aaacaatctt 300
109 cttcgagggt ggtggagaaa ttgcccggt agttgactac agaccacagg aagacggaac 360
111 tgagaaaact tttacaagaa aattctctag caaatgccca ggcaattaca tgcttatgga 420
113 cgtgtgcgct acaagggacg ctgatgataa atgcacgaa ggcacaattg tggtgacagt 480
115 caggggtgcc ctatatgacg aagataacaa tgggtgaatg gatgaaggta aggttattcc 540
117 atctgagaca atcgaggatg atatcaagga ctgtgggctc ttagaccaag atgttgaact 600
119 cgattatacg tggactcaaa acgagtgtga tctaccagac acagtagacg aggctgaaga 660
121 cacaccgtca gaaactggag aattcttctg gtagatctat cagaccactt ttatcagcag 720
123 gacaactggg cgttaccaga cacctataac gtgtcctcat caataatgtg taaaacagaa 780
125 ataatcgata gaattattgaa aataaaatgt taatagacac tggttgaaaa aaaaaaaaaa 840
127 aaaaaactcg ag 852
131 <210> SEQ ID NO: 4
132 <211> LENGTH: 225
133 <212> TYPE: PRT
134 <213> ORGANISM: Pholas dactylus
136 <400> SEQUENCE: 4
137 Met Ala Cys Ile Val Phe Val Ala Leu Val Ala Leu Cys Leu Met Gln
138 1 5 10 15
140 Pro Gly Ser Gly Glu Glu Val Gln Cys Ala Met Asn Trp Thr Gln Ala
141 20 25 30
143 Asn Glu Tyr Val Phe Asn Val Asp Trp Met Thr Ile Phe Ile Tyr Asp
144 35 40 45
146 Tyr Gly Ala Gln Glu Gln Leu Tyr Glu Asp Arg Ala Leu Gly Leu Cys
147 50 55 60
149 Arg Ile Glu Arg Ala Gly Pro Gly Thr Thr Lys Ala Val Trp Ile Asn
151 65 70 75 80
153 Trp Ser Asn Asp Thr Gln Ser Cys Val Thr Arg Lys Thr Ile Phe Phe
154 85 90 95
156 Glu Val Gly Gly Glu Ile Ala Arg Leu Val Asp Tyr Arg Pro Gln Glu
157 100 105 110
159 Asp Gly Thr Glu Lys Thr Phe Thr Arg Lys Phe Ser Ser Lys Met Pro
160 115 120 125
162 Gly Thr Tyr Met Leu Met Asp Val Cys Ala Thr Arg Asp Ala Asp Asp
163 130 135 140
165 Lys Cys Ile Glu Gly Thr Ile Val Val Thr Val Arg Val Ser Leu Tyr
166 145 150 155 160
168 Asp Glu Asp Asn Asn Gly Val Met Asp Glu Gly Lys Val Ile Pro Ser

```

RAW SEQUENCE LISTING

DATE: 03/15/2002

PATENT APPLICATION: US/09/831,142A

TIME: 11:08:09

Input Set : A:\es.txt

Output Set: N:\CRF3\03152002\I831142A.raw

```

169          165          170          175
171 Glu Thr Ile Glu Asp Asp Ile Lys Asp Cys Gly Leu Leu Asp Gln Asp
172          180          185          190
174 Val Glu Leu Asp Tyr Thr Trp Thr Gln Asn Glu Cys Asp Leu Pro Asp
175          195          200          205
177 Thr Val Asp Glu Ala Glu Asp Thr Pro Ser Glu Thr Gly Glu Phe Phe
178          210          215          220
180 Trp
181 225
184 <210> SEQ ID NO: 5
185 <211> LENGTH: 205
186 <212> TYPE: PRT
187 <213> ORGANISM: Pholas dactylus
189 <400> SEQUENCE: 5
190 Glu Glu Val Gln Cys Ala Met Asn Trp Thr Gln Ala Asn Glu Tyr Val
191 1 5 10 15
193 Phe Asn Val Asp Trp Met Thr Ile Phe Ile Tyr Asp Tyr Gly Ala Gln
194 20 25 30
196 Glu Gln Leu Tyr Glu Asp Arg Ala Leu Gly Leu Cys Arg Ile Glu Arg
197 35 40 45
199 Ala Gly Pro Gly Thr Thr Lys Ala Val Trp Ile Asn Trp Ser Asn Asp
201 50 55 60
203 Thr Gln Ser Cys Val Thr Arg Lys Thr Ile Phe Phe Glu Val Gly Gly
204 65 70 75 80
206 Glu Ile Ala Arg Leu Val Asp Tyr Arg Pro Gln Glu Asp Gly Thr Glu
207 85 90 95
209 Lys Thr Phe Thr Arg Lys Phe Ser Ser Lys Met Pro Gly Thr Tyr Met
210 100 105 110
212 Leu Met Asp Val Cys Ala Thr Arg Asp Ala Asp Asp Lys Cys Ile Glu
213 115 120 125
215 Gly Thr Ile Val Val Thr Val Arg Val Ser Leu Tyr Asp Glu Asp Asn
216 130 135 140
218 Asn Gly Val Met Asp Glu Gly Lys Val Ile Pro Ser Glu Thr Ile Glu
219 145 150 155 160
221 Asp Asp Ile Lys Asp Cys Gly Leu Leu Asp Gln Asp Val Glu Leu Asp
222 165 170 175
224 Tyr Thr Trp Thr Gln Asn Glu Cys Asp Leu Pro Asp Thr Val Asp Glu
225 180 185 190
227 Ala Glu Asp Thr Pro Ser Glu Thr Gly Glu Phe Phe Trp
228 195 200 205
231 <210> SEQ ID NO: 6
232 <211> LENGTH: 225
233 <212> TYPE: PRT
234 <213> ORGANISM: Pholas dactylus
236 <400> SEQUENCE: 6
237 Met Ala Cys Ile Val Phe Val Ala Leu Val Ala Leu Cys Leu Met Gln
238 1 5 10 15
240 Pro Gly Ser Gly Glu Glu Val Gln Cys Ala Met Asn Trp Thr Gln Ala
241 20 25 30

```

RAW SEQUENCE LISTING

DATE: 03/15/2002

PATENT APPLICATION: US/09/831,142A

TIME: 11:08:09

Input Set : A:\es.txt

Output Set: N:\CRF3\03152002\I831142A.raw

```

243 Asn Glu Tyr Val Phe Asn Val Asp Trp Met Thr Ile Phe Ile Tyr Asp
244           35                      40                      45
246 Tyr Gly Ala Gln Glu Gln Leu Tyr Glu Asp Arg Ala Leu Gly Leu Cys
247           50                      55                      60
249 Arg Ile Glu Arg Ala Gly Pro Gly Thr Thr Lys Ala Val Trp Ile Asn
251 65                      70                      75                      80
253 Trp Ser Asn Asp Thr Gln Ser Cys Val Thr Arg Lys Thr Ile Phe Phe
254                      85                      90                      95
256 Glu Val Gly Gly Glu Ile Ala Arg Leu Val Asp Tyr Arg Pro Gln Glu
257           100                      105                      110
259 Asp Gly Thr Glu Lys Thr Phe Thr Arg Lys Phe Ser Ser Lys Met Pro
260           115                      120                      125
262 Gly Thr Tyr Met Leu Met Asp Val Cys Ala Thr Arg Asp Ala Asp Asp
263           130                      135                      140
265 Lys Cys Ile Glu Gly Thr Ile Val Val Thr Val Arg Val Ser Leu Tyr
266 145                      150                      155                      160
268 Asp Glu Asp Asn Asn Gly Val Met Asp Glu Gly Lys Val Ile Pro Ser
269           165                      170                      175
271 Glu Thr Ile Glu Asp Asp Ile Lys Asp Cys Gly Leu Leu Asp Gln Asp
272           180                      185                      190
274 Val Glu Leu Asp Tyr Thr Trp Thr Gln Asn Glu Cys Asp Leu Pro Asp
275           195                      200                      205
277 Thr Val Asp Glu Ala Glu Asp Thr Pro Ser Glu Thr Gly Glu Phe Phe
278           210                      215                      220

```

280 Trp

281 225

284 <210> SEQ ID NO: 7

285 <211> LENGTH: 17

286 <212> TYPE: DNA

287 <213> ORGANISM: Pholas dactylus

289 <220> FEATURE:

290 <221> NAME/KEY: modified_base

291 <222> LOCATION: (3)

292 <223> OTHER INFORMATION: i

294 <400> SEQUENCE: 7

W--> 295 acnathhtyt tycargt

17

299 <210> SEQ ID NO: 8

301 <211> LENGTH: 17

302 <212> TYPE: DNA

303 <213> ORGANISM: Pholas dactylus

305 <220> FEATURE:

306 <221> NAME/KEY: modified_base

307 <222> LOCATION: (15)

308 <223> OTHER INFORMATION: i

310 <400> SEQUENCE: 8

W--> 311 cargargarg gnacnga

17

315 <210> SEQ ID NO: 9

316 <211> LENGTH: 17

317 <212> TYPE: DNA

RAW SEQUENCE LISTING

DATE: 03/15/2002

PATENT APPLICATION: US/09/831,142A

TIME: 11:08:09

Input Set : A:\es.txt

Output Set: N:\CRF3\03152002\I831142A.raw

```

318 <213> ORGANISM: Pholas dactylus
320 <220> FEATURE:
321 <221> NAME/KEY: modified_base
322 <222> LOCATION: (3)
323 <223> OTHER INFORMATION: i
325 <400> SEQUENCE: 9
W--> 326 tcngtnccyt cytcytg 17
330 <210> SEQ ID NO: 10
331 <211> LENGTH: 18
332 <212> TYPE: DNA
333 <213> ORGANISM: Pholas dactylus
335 <220> FEATURE:
336 <221> NAME/KEY: modified_base
337 <222> LOCATION: (9)
338 <223> OTHER INFORMATION: i
340 <400> SEQUENCE: 10
W--> 341 ttayaagtng aytggatg 18
345 <210> SEQ ID NO: 11
346 <211> LENGTH: 20
347 <212> TYPE: DNA
348 <213> ORGANISM: Pholas dactylus
351 <400> SEQUENCE: 11
352 acacagcccc aaagcccgat 20
356 <210> SEQ ID NO: 12
357 <211> LENGTH: 20
358 <212> TYPE: DNA
359 <213> ORGANISM: Pholas dactylus
361 <400> SEQUENCE: 12
362 ttgcccggct agttgactac 20
366 <210> SEQ ID NO: 13
367 <211> LENGTH: 24
368 <212> TYPE: DNA
369 <213> ORGANISM: Pholas dactylus
371 <400> SEQUENCE: 13
372 catatttcaa ccagtgttta ttaa 24
376 <210> SEQ ID NO: 14
377 <211> LENGTH: 19
378 <212> TYPE: DNA
379 <213> ORGANISM: Pholas dactylus
381 <400> SEQUENCE: 14
382 caattgtgcc ttcgatgca 19
386 <210> SEQ ID NO: 15
387 <211> LENGTH: 17
388 <212> TYPE: DNA
389 <213> ORGANISM: Pholas dactylus
391 <400> SEQUENCE: 15
392 ggactgtggg ctcttag 17
396 <210> SEQ ID NO: 16
397 <211> LENGTH: 20

```

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 03/15/2002
PATENT APPLICATION: US/09/831,142A TIME: 11:08:10

Input Set : A:\es.txt
Output Set: N:\CRF3\03152002\I831142A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:7; N Pos. 3
Seq#:8; N Pos. 12,15
Seq#:9; N Pos. 3,6
Seq#:10; N Pos. 9

VERIFICATION SUMMARY

DATE: 03/15/2002

PATENT APPLICATION: US/09/831,142A

TIME: 11:08:10

Input Set : A:\es.txt

Output Set: N:\CRF3\03152002\I831142A.raw

L:10 M:270 C: Current Application Number differs, Replaced Current Application Number
L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:295 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:7 after pos.:0
L:311 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:8 after pos.:0
L:326 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:9 after pos.:0
L:341 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:10 after pos.:0